Stat120B: Introduction to Prabability and Statistics **Lecture Notes Archive**

Cramer-Rao Inequality

Lecturer: Gang Liang

Cramer-Rao Inequality gives an information lower bound that an estimate can possibly obtain.

Theorem 1 *Let $X_1, \cdots, X_n$ be i.i.d. sample with density function $f(x;\theta)$. Let $T = t(X_1, \cdots, X_n)$ be an unbiased estimate of unknown parameter θ. Under some general smoothness assumption of $f(x;\theta)$, we have*

$$\mathrm{Var}(T) \geq \frac{1}{nI(\theta)}.$$

Proof: We will prove this theorem in three steps. First, the log-likelihood function of the problem is

$$Z = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i;\theta).$$

We would like to show that its expectation is just zero. The log-likelihood function Z is the summation of many i.i.d. terms, so we only need to compute the expectation of one term.

$$\begin{aligned} \mathrm{E}\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right) &= \int \left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) f(x;\theta)\mathrm{d}x = \int \frac{\partial}{\partial\theta} f(x;\theta)\mathrm{d}x \\ &= \frac{\partial}{\partial\theta}\left(\int f(x;\theta)\mathrm{d}x\right) \text{ (Note that } \textstyle\int f(x;\theta)\mathrm{d}x = 1) \\ &= 0. \end{aligned}$$

Second, similarly, we can show $\mathrm{Var}(Z) = nI(\theta)$. Now let us consider the covariance/correlation of Z and T.

$$\frac{\mathrm{Cov}^2(Z,T)}{\mathrm{Var}(Z)\mathrm{Var}(T)} = \mathrm{Corr}^2(Z,T) \leq 1,$$

i.e.,

$$\mathrm{Var}(T) \geq \frac{\mathrm{Cov}^2(Z,T)}{\mathrm{Var}(Z)} = \frac{\mathrm{Cov}^2(Z,T)}{nI(\theta)}.$$

Third, we want to show that the $\mathrm{Cov}(Z,T)$ is equal to 1. Easy to see that it will finish the proof. Notice that $\mathrm{E}Z = 0$,

$$\mathrm{Cov}(Z,T) = \mathrm{E}\left((Z - \mathrm{E}(Z))(T - \mathrm{E}(T))\right) = \mathrm{E}(ZT) - \mathrm{E}(Z)\mathrm{E}(T) = \mathrm{E}(ZT).$$

Now we have

$$\begin{aligned}
\mathrm{Cov}(Z,T) &= \int\cdots\int t(x_1,\cdots,x_n)\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right]\prod_{i=1}^{n} f(x_i;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \int\cdots\int t(x_1,\cdots,x_n)\left[\sum_{i=1}^{n}\frac{\frac{\partial}{\partial\theta}f(x_i;\theta)}{f(x_i;\theta)}\right]\prod_{i=1}^{n} f(x_i;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \frac{\partial}{\partial\theta}\left(\int\cdots\int t(x_1,\cdots,x_n)\prod_{i=1}^{n} f(x_i;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n\right) \\
&= \frac{\partial}{\partial\theta}\mathrm{E}(T) = \frac{\partial}{\partial\theta}\theta \\
&= 1.
\end{aligned}$$

Some notes on the CR lower bound:

- The asymptotic variance of the MLE is $nI(\theta)$, and this theorem further shows that it is also the lower bound that any unbiased estimator can possibly achieve, so it implies that asymptotically, the MLE is the optimal estimator based on the variance criteria.
- The lower bound provided by the theorem is very conservative. In many cases, it is not achieved. Also it is important to note that the MLE achieves this bound asymptotically. It does not imply the MLE will reach the bound in any finite sample examples.
- More generally, if T is an unbiased estimator of $\tau(\theta)$, then we have

$$\mathrm{Var}(T) \geq \frac{\tau'^2(\theta)}{nI(\theta)}.$$